The Influence of Visual Cues on Sound Externalization

Carvajal, Juan Camilo Gil; Santurette, Sébastien; Cubick, Jens; Dau, Torsten

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

The Influence of Visual Cues on Sound Externalization

Juan Camilo Gil Carvajal, Sébastien Santurette, Jens Cubick, Torsten Dau

Hearing Systems, Department of Electrical Engineering, Technical University of Denmark

Introduction

Sounds presented via headphones are typically perceived inside the head. However, the illusion of a sound source located out in space away from the listener's head can be generated with binaural headphone-based auralization systems by convolving anechoic sound signals with a binaural room impulse response (BRIR) measured with miniature microphones placed in the listener's ear canals. Sound externalization of such virtual sounds can be very convincing and robust but there have been reports that the illusion might break down when the listening environment differs from the room in which the BRIRs were recorded [1,2,3]. This may be due to incongruent auditory cues between the recording and playback room during sound reproduction [2]. Alternatively, an expectation effect caused by the visual impression of the room may affect the position of the perceived auditory image [3]. Here, we systematically investigated whether incongruent auditory and visual room-related cues affected sound externalization in terms of perceived distance, azimuthal localization, and compactness.

Methods

- Three playback rooms were used:
  - **IEC$_M$:** Medium-sized IEC standard ($V \approx 100$ m^3, $T_{60} = 0.4$ s) in which all BRIRs were recorded
  - **Rev$_S$:** Small reverberant ($V \approx 43$ m^3, $T_{60} = 2.8$ s)
  - **Dry$_L$**: Large anechoic ($V \approx 330$ m^3, $T_{60} < 0.01$ s)
- Eighteen naïve listeners were divided into two groups (Fig.1):
  - **A:** Blindfolded during testing but with *auditory awareness* of the room provided by an in-room noise source
  - **V:** Shielded from room-related acoustic input but with *visual awareness* of the room
- All listeners were also tested with all cues (**VA**) available.

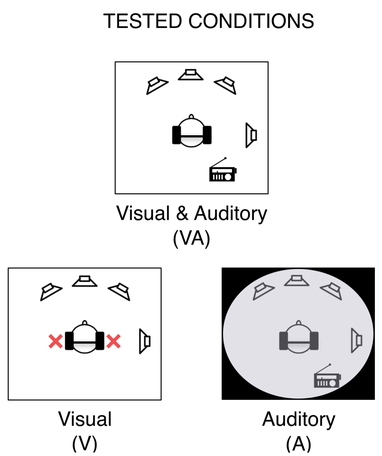

Fig.1 Overview of tested conditions.

- Seven azimuthal positions (II, III, VI, VII, IX, XI, and XII o'clock) were reproduced (Fig.2).
- Loudspeakers visible at positions I, III, XI, and XII.
- Subjective rating scales for distance, azimuth, and compactness perception (Fig.2).

Experimental setup

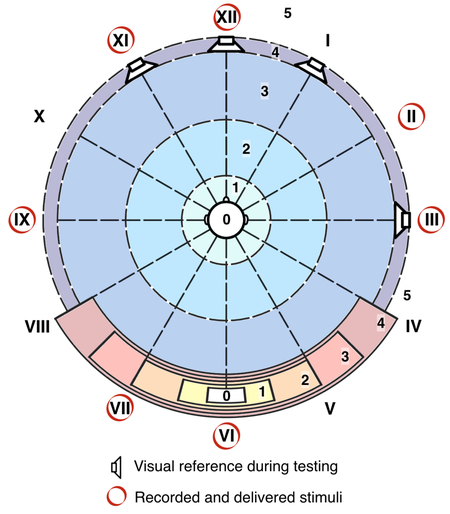

Fig.2 Loudspeaker setup and subjective rating scales used for the experiments. For distance judgments, listeners could provide ratings of 0 (inside the head), 1, 2, 3, 4 (at the loudspeaker), or 5. For compactness judgments, listeners could provide ratings of 0 (most compact), 1, 2, 3, 4, or 5 (least compact). The red circles indicate the reproduced azimuthal directions in the experiment. Visible loudspeaker positions are indicated by the loudspeaker pictograms.

Results: Distance

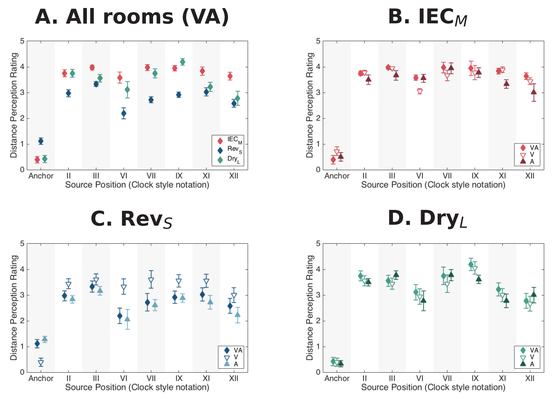

Fig.3 Distance perception ratings obtained in the three rooms. **A**: Condition VA across rooms. **B**: Conditions VA, A and V in the IEC$_M$ room. **C**: Conditions VA, A and V in the Rev$_S$ room. **D**: Conditions VA, A and V in the Dry$_L$ room.

- Significant effect of playback room: lower VA ratings in Rev$_S$ for all positions, in Dry$_L$ for positions III, VI, XI, and XII.
- In Rev$_S$, no difference between A and VA conditions, but significantly higher ratings for V than VA for most positions.
- In Dry$_L$, no significant differences between A, V, and VA conditions for most positions.

Results: Azimuth

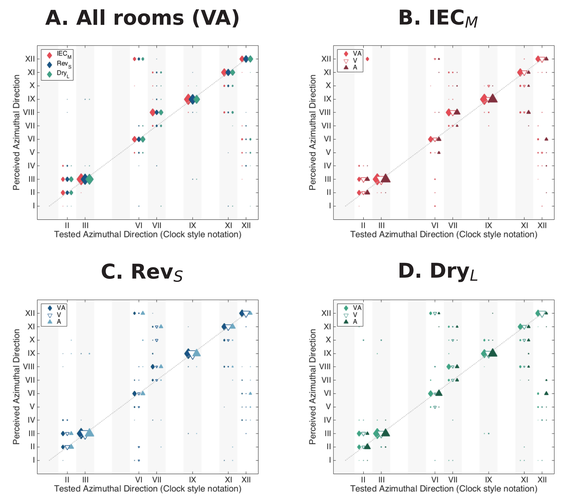

Fig.4 Azimuthal direction ratings obtained in the three rooms. **A**: Condition VA across rooms. **B**: Conditions VA, A and V in the IEC$_M$ room. **C**: Conditions VA, A and V in the Rev$_S$ room. **D**: Conditions VA, A and V in the Dry$_L$ room.

- No significant effect of playback room (similar confusions).
- Visual capturing effect of loudspeaker at III for sounds presented at II, with a stronger effect for V vs VA and VA vs A.
- Less front-back confusions for VA vs A.

Results: Compactness

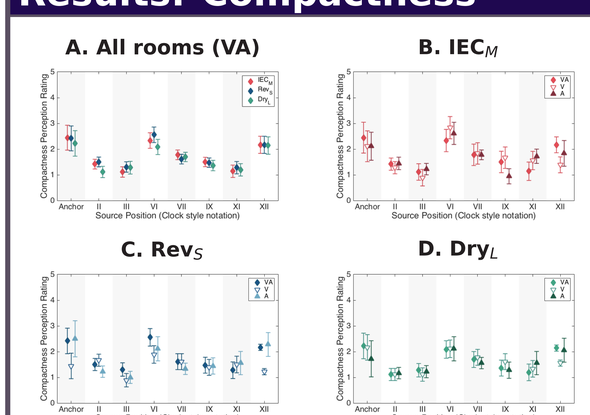

Fig.5 Compactness ratings obtained in the three rooms. **A**: Condition VA across rooms. **B**: Conditions VA, A and V in the IEC$_M$ room. **C**: Conditions VA, A and V in the Rev$_S$ room. **D**: Conditions VA, A and V in the Dry$_L$ room.

- No significant effect of playback room.
- Similar but less consistent ratings when cues from either modality are limited.
- Auditory images consistently more compact for lateral positions than for front and back positions.

Externalization scores

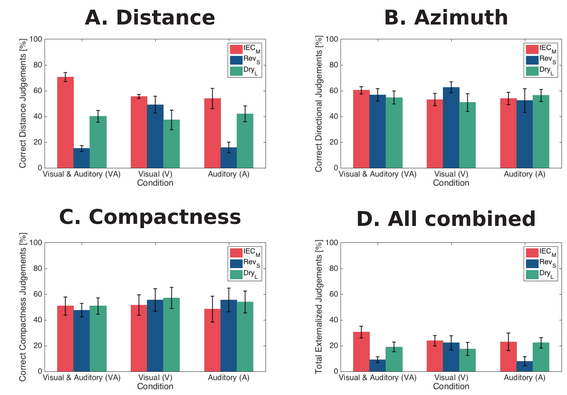

Fig.6 Total percentages of correct judgements in every condition per room. **A**: Correct distance ratings. **B**: Correct azimuthal direction ratings. **C**: Correct compactness ratings. **D**: Combined externalization.

- Correct overall externalization ratings defined as those with a score of 4 for distance, localized at the correct azimuthal direction, and within the range 0 to 1 for compactness.
- Percent correct scores pooled across positions confirm that the effects of playback room and reduced auditory or visual room-related input only affect distance perception but not azimuthal localization and compactness.
- Low combined scores reflect the fact that the percentages of correct judgments do not necessarily covary across attributes and differ across positions.

Conclusions

- A mismatch between recording and playback room is detrimental to virtual sound externalization and affects mainly the perceived distance but not the azimuthal localization and compactness of the auditory image.
- The auditory modality governs externalization in terms of perceived distance when cues from the recording and playback room are incongruent.
- In incongruent listening situations, the more reverberant the playback room is, the more critical the auditory impression of the room seems to become.
- The visual impression of the room does not affect perceived distance, but source-related visual cues help resolve localization ambiguities and improve compactness perception.

References:

[1] Begault DR (**1999**). "Auditory and non-auditory factors that potentially influence virtual acoustic imagery," Proc. Audio Eng. Soc. **16**:13-26.
[2] Werner S, Siegel A (**2011**). "Effects of binaural auralization via headphones on the perception of acoustic scenes," Proc. ISAAR **3**:215-222.
[3] Udesen J, Piechowiak T, Gran F (**2015**). "The effect of vision on psychoacoustic testing with headphone-based virtual sound," J. Audio Eng. Soc. **63**:552-561.

Presented at: 39th MidWinter Meeting of the Association for Research in Otolaryngology — February 20-24, 2016, San Diego, CA — Poster #PS-217.

Contact: Juan Camilo Gil Carvajal — *s130655@student.dtu.dk*
Sébastien Santurette — *ses@elektro.dtu.dk*
Jens Cubick — *jecu@elektro.dtu.dk*
Torsten Dau — *tdau@elektro.dtu.dk*